

# A Poisson model for earthquake frequency uncertainties in seismic hazard analysis

J. Greenhough & I. G. Main

School of Geosciences, University of Edinburgh, UK  
john.greenhough@ed.ac.uk, ian.main@ed.ac.uk

September 3, 2008

## Abstract

Frequency-magnitude distributions, and their associated uncertainties, are of key importance in statistical seismology. When fitting these distributions, the assumption of Gaussian residuals is invalid since event numbers are both discrete and of unequal variance. In general, the observed number in any given magnitude range is described by a binomial distribution which, given a large total number of events of all magnitudes, approximates to a Poisson distribution for a sufficiently small probability associated with that range. In this paper, we examine four earthquake catalogues: New Zealand (Institute of Geological and Nuclear Sciences), Southern California (Southern California Earthquake Center), the Preliminary Determination of Epicentres and the Harvard Centroid Moment Tensor (both held by the United States Geological Survey). Using independent Poisson distributions to model the observations, we demonstrate a simple way of estimating the uncertainty on the total number of events occurring in a fixed time period.

## 1 Introduction

It is well documented that typical catalogues containing large numbers of earthquake magnitudes are closely approximated by power-law or gamma frequency distributions [1, 2, 3, 4]. This paper addresses the characterisation of counting errors (that is, the uncertainties in histogram frequencies) required when fitting such a distribution via the maximum likelihood method, rather than the choice of model itself (for which see [5]). We follow this with an empirical demonstration of the Poisson approximation for total event-rate uncertainty [used in 5]. Our analysis provides evidence to support the assumption in seismic hazard assessment that earthquakes are Poisson processes [6, 7, 8, 9], which is routinely stated yet seldom tested or used as a constraint when fitting frequency-magnitude distributions. Use is made of the Statistical Seismology Library [10],

specifically the data downloaded from the New Zealand Institute of Geological and Nuclear Sciences (GNS, <http://www.gns.cri.nz>), the Southern California Earthquake Center (SCEC, <http://www.scec.org>) and the United States Geological Survey (USGS, <http://www.usgs.gov>), along with associated R functions for extracting the data.

Consider a large sample of  $N$  earthquakes. In order to estimate the underlying proportions of different magnitudes, which reflect physical properties of the system, the data are binned into  $m$  magnitude ranges containing  $\mathbf{n}$  events such that  $\sum_{i=1}^m n_i = N$ . Since  $\mathbf{n}$  are discrete, a Gaussian model for each  $n_i$  is inappropriate and may introduce significant biases in parameter estimations [11, 12, 13]. Hence when fitting some relationship with magnitudes  $\mathbf{M}$ ,  $\mathbf{n}_{fit} = f(\mathbf{M})$ , linear regression must take the generalised, rather than least-squares, form [14]. Weighted least squares is an alternative approach which we do not consider here. The set  $\mathbf{n}$  is described as a multinomial distribution; should we wish to test whether two different samples  $\mathbf{n}$  and  $\mathbf{n}'$  are significantly different given a fixed  $N$  “trials”, confidence intervals that reflect the simultaneous occurrence of all  $\mathbf{n}$  must be constructed using a Bayesian approach [15]. However, in the case of earthquake catalogues, it is the temporal duration rather than the number of events that is fixed. Observational variability is not, therefore, constrained to balance a higher  $n_i$  at some magnitude with a lower  $n_j$  elsewhere, and  $\mathbf{n}$  are well approximated by independent binomial distributions [16].

Each incremental magnitude range  $(M_i - \delta M/2, M_i + \delta M/2)$  contains a proportion of the total number of events and hence a probability  $p_i$  with which any event will fall in that range. Providing the overall duration of the catalogue is greater than that of any significant correlations between either magnitudes or inter-event times,  $n_i$  can be modelled as a binomial experiment with  $N$  independent trials each having a probability of “success”  $p_i$  [16]. The binomial distribution converges towards the Poisson distribution as  $N \rightarrow \infty$  while  $Np_i$  remains fixed. Various rules of thumb are quoted to suggest values of  $N$  and  $p_i$  for which a Poisson approximation may be valid; see for example [17, 18]. Here, we show empirically in Sect. 2 that the frequencies in four natural earthquake catalogues are consistent with a Poisson hypothesis, while in Sect. 3 we derive the resulting Poisson distributions of the total numbers of events, which provide simple measures of uncertainty in event rates.

## 2 Frequency-magnitude Distributions

Four earthquake catalogues are analysed: New Zealand (1460 – Mar 2007), Southern California (Jan 1932 – May 2007), the Preliminary Determination of Epicentres (PDE, Jan 1964 – Sep 2006) and the Harvard Centroid Moment Tensor (CMT, Jan 1977 – June 1999, <100 km focal depth). While we impose no additional temporal or spatial filters on the raw data, magnitude limits are chosen to minimise the effects of incompleteness at lower magnitudes and undersampling of higher magnitudes. Following [5], who demonstrate the use of an objective Bayesian information criterion for choosing between functions, we

seek to fit each catalogue with either a single power-law distribution

$$\log_{10} \mathbf{n} = a - b\mathbf{M}, \quad (1)$$

$\mathbf{M}$  being already on a log scale, or a gamma distribution

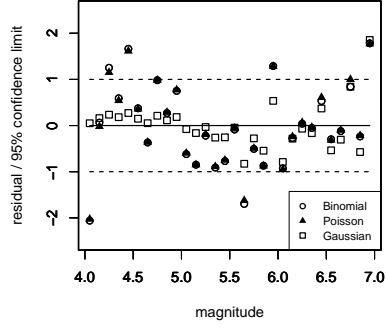
$$\log_{10} \mathbf{n} = a - b\mathbf{M} - c \exp(k\mathbf{M}), \quad (2)$$

where  $a$ ,  $b$ ,  $c$  and  $k$  are constants. The gamma distribution consists of a power law of seismic moment or energy at the lower magnitudes followed by an exponential roll-off. Unlike pure power laws, its integration is finite and so it represents a physical generalisation of the Gutenberg-Richter law; for examples see [19] and references therein. For internal consistency, the Poisson assumption in [5] is indeed valid as we now demonstrate.

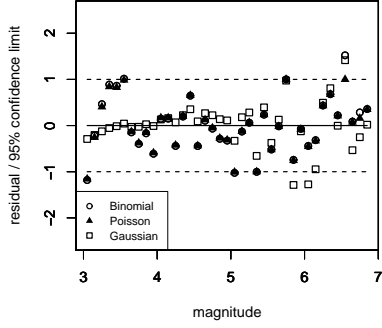
As explained in Sect. 1, generalised linear regression is required since we have non-Gaussian counting errors on each bin. To test the consistency of these counting errors with the Gaussian, binomial and Poisson distributions, the residuals (observations minus chosen fit) are normalised to their 95% confidence intervals and plotted in Fig. 1. In all four catalogues, the binomial and Poisson residuals are almost indistinguishable, and show no significant deviation from the expected 1 in 20 exceedance rate when counting those points that lie outside the 95% confidence limits. Equal bin widths  $\Delta M = 0.1$  are used as is common practice in earthquake hazard analysis; while this underestimates the intrinsic physical uncertainty of earthquake magnitude determination, for the present purposes the Poisson model appears to be a good proxy. At least, for the catalogues considered here and with  $\Delta M = 0.1$ , the Poisson model is valid. By way of a further check, the value  $b$  of the fitted power-law slope (Eqs. 1, 2) given binomial errors is, to two significant figures, equal to that given Poisson errors, for all four catalogues. Constant Gaussian errors systematically overestimate frequency uncertainties on the smaller magnitudes, leading to differences in  $b$  of +10% and -30% respectively for the Southern California and PDE data (see caption of Fig. 2). These are caused by over-weighting the exponential components of the gamma distributions and exemplify worst-case results of incorrect error structures. In Fig. 2, then, we need only plot the fits and uncertainties using the Poisson model. Let us now describe, in Sect. 3, the usefulness of this result for estimating event-rate uncertainties.

### 3 Event-rate Uncertainties

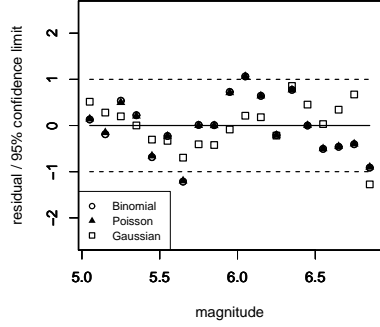
Having established that independent Poisson distributions characterise the magnitude frequencies in these four catalogues (importantly, these data span sufficiently large times and distances as to minimise dependencies due to clustering), we now ask how this impacts on uncertainties in total numbers of events. While we cannot create equivalent catalogues by re-sampling the same regions under the same physical conditions, we can simulate  $S = 10^5$  samples from each magnitude range by keeping the fitted mean  $\lambda_i$  constant (representing the underlying



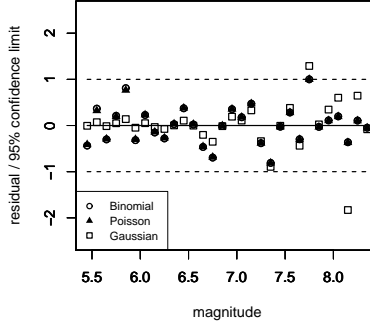
(a)



(b)



(c)



(d)

Figure 1: Residuals of fitted frequency-magnitude distributions from GNS/SCEC/USGS catalogues: (a) New Zealand, (b) Southern California, (c) PDE, (d) CMT. (solid line) Best fit to Eq. 1 or 2; (dashed lines) 95% confidence limits of respective distribution.

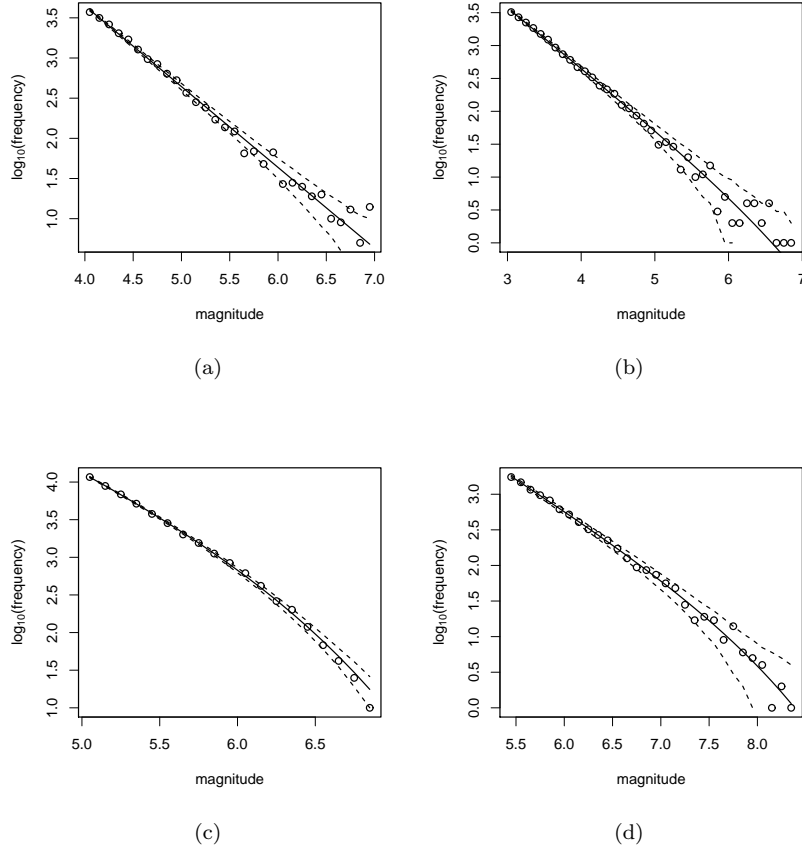


Figure 2: Frequency-magnitude distributions from GNS/SCEC/USGS catalogues. (solid line) Best fit to Eq. 1 or 2: (a) New Zealand, power law  $b = 1.0$ ; (b) Southern California, gamma  $b = 0.91$ ; (c) PDE, gamma  $b = 0.91$ ; (d) CMT, gamma  $b = 0.85$ . (dashed lines) 95% Poisson confidence limits. Unweighted Gaussian regression leads to  $b$ -value estimates of (a) 0.98, (b) 1.03, (c) 0.66, (d) 0.83.

reality) and using the Poisson estimate  $\sigma_i^2 = \lambda_i$  to capture the observational variance. Summing these realisations, one per bin over all magnitudes, provides a large set of plausible alternative totals. Figure 3 shows histograms of these simulated totals for each of the four catalogues, fitted with Poisson distributions for reasons we now explain.

It is straightforward to show analytically that the sum of independent Poisson variables is itself Poisson with a mean (and hence variance) equal to the sum of the component means  $\lambda$  [16]. This result holds for (i) any number of independent Poisson variables (in the current context, bins) with (ii) any relationship  $\lambda = f(\mathbf{M})$ , since the result is independent of  $f(\mathbf{M})$ . In the case of earthquakes placed into bins of width  $\Delta M$  at magnitudes  $\mathbf{M}$ , for example,  $f(\mathbf{M})$  is commonly fitted by a power-law or gamma distribution as in Fig. 2. From the Poisson property  $\sigma^2 = \lambda$ , it follows that

$$\sigma_N^2 = \lambda_N = \sum \lambda = \sum f(\mathbf{M}). \quad (3)$$

Thus we have a useful result: if there exists a physically justifiable function that provides a satisfactory fit to the histogram (that is, Poisson-distributed uncorrelated residuals as in Fig. 1) then the mean and variance of the total number of events, over different realisations of the catalogue, are both equal to the sum of the fitted values (Eq. 3). For the simulations of our four example catalogues (Fig. 3), we have mean total event numbers of  $\lambda_N = 19231, 17491, 46454, 9301$  respectively; these match the actual observed totals to an accuracy of  $\pm 1$ . Empirical evaluations confirm  $\sigma_N = \sqrt{\lambda_N}$  to two significant figures, hence our estimated uncertainties on total event numbers for these catalogues are  $\sigma_N = 140, 130, 220, 96$ . Since (i) a Poisson distribution converges towards a Gaussian as  $\lambda \rightarrow \infty$ , (ii) a reasonable approximation to this exists where  $\lambda > 5$  and  $S - \lambda > 5$  for sample size  $S$  [20], and (iii) we have  $S = 10^5$  with  $\lambda_N$  given above, it is not surprising that the Poisson confidence intervals for  $\lambda_N \pm \sigma_N$  are (to two significant figures) 68% as in the Gaussian case.

## 4 Conclusions

The purpose of this paper is to draw attention to the simplicity with which one can formally estimate event-rate uncertainties for applications in seismic hazard analysis, both in small magnitude ranges and over whole catalogues. For each of the four earthquake catalogues considered here, we find that the best estimate of both the mean and the variance of the total number of events, is equal to the total calculated from the fit to the histogram. This approximation holds where (i) the residuals of the fit are independently Poisson distributed, and (ii) the overall duration of the catalogue is greater than that of any significant correlations between either magnitudes or inter-event times. Note that the ratio of binomial-to-Poisson variance for any frequency  $n$  is  $\sigma_b^2/\sigma_P^2 = 1 - p_n < 1$ , which implies that the Poisson approximation provides an upper bound for the uncertainty on the total event rate should any residuals generalise to the

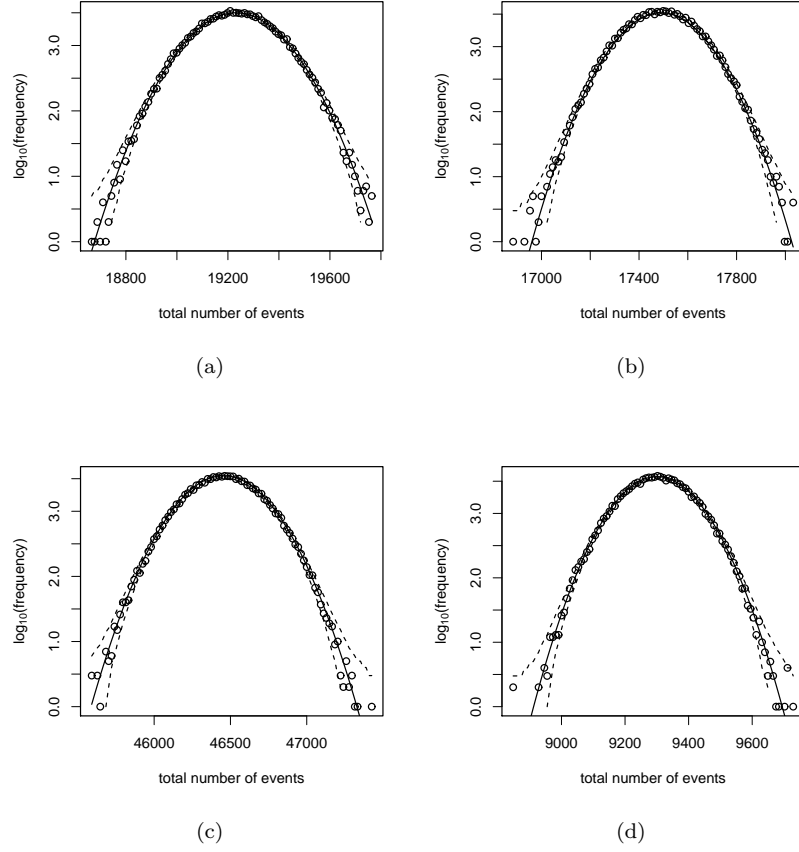


Figure 3: Event-rate distributions from  $10^5$  simulated realisations of GNS/SCEC/USGS catalogues. Each total event-rate is the sum of a random sample of frequencies, one per bin, given Poisson uncertainties shown in Fig. 2. (a) New Zealand, (b) Southern California, (c) PDE, (d) CMT. (solid line) Best fit Poisson distribution; (dashed lines) 99% binomial confidence limits.

binomial case. However, correlations between inter-event times could cause significant future changes in event rates, greater than predicted by the naive estimates of uncertainty presented here, and this is the subject of further study.

## Acknowledgements

The authors gratefully acknowledge financial support from the New and Emerging Science and Technologies Pathfinder program “Triggering Instabilities in Materials and Geosystems” (contract NEST-2005-PATH-COM-043386).

## References

- [1] Richter, C. H. (1958), *Elementary Seismology*, W. H. Freeman, New York.
- [2] Turcotte, D. L. (1992), *Fractals and Chaos in Geology and Geophysics*, Cambridge Univ. Press, New York.
- [3] Main, I. G. (1996), Statistical physics, seismogenesis, and seismic hazard, *Rev. Geophys.*, *34*, 433–462.
- [4] Main, I. G., L. Li, J. McCloskey and M. Naylor (2008), Effect of the Sumatran mega-earthquake on the global magnitude cut-off and event rate, *Nature Geosci.*, *1*, 142, doi:10.1038/ngeo141.
- [5] Leonard, T., O. Papasouliotis and I. G. Main (2001), A Poisson model for identifying characteristic size effects in frequency data: Application to frequency-size distributions for global earthquakes, “starquakes”, and fault lengths, *JGR*, *106*, B7, 13473–13484.
- [6] Reiter, L. (1990), *Earthquake Hazard Analysis: Issues and Insights*, Columbia Univ. Press, New York.
- [7] Bozorgnia, Y. and V. V. Bertero (2004), *Earthquake Engineering: From Engineering Seismology to Performance-Based Engineering*, CRC Press, Boca Raton, USA.
- [8] Lombardi, A. M., A. Akinci, L. Malagnini and C. S. Mueller (2005), Uncertainty analysis for seismic hazard in Northern and Central Italy, *Ann. Geophys.*, *48*, 853–865.
- [9] Kossobokov, V. G. (2006), Testing earthquake prediction methods: The West Pacific short-term forecast of earthquakes with magnitude  $M_w \geq 5.8$ , *Tectonophysics*, *413*, 25–31.
- [10] Harte, D., Statistical Seismology Library, <http://homepages.paradise.net.nz/david.harte/SSLib>, Statistics Research Associates, Wellington, New Zealand.



- [11] Aki, K. (1965), Maximum likelihood estimation of  $b$  in the formula  $\log N = a - bm$  and its confidence limits, *Bull. Earthquake Res. Inst. Univ. Tokyo*, 43, 237–239.
- [12] Keilis-Borok, V. I., L. V. Kantorovich, Ye. V. Vilkovich and G. M. Molchan (1970), Statistical seismicity model and estimation of principal seismic effects, *Izvestia AN SSSR. Phys. Earth*, 5, 85–101 (in Russian).
- [13] Sandri, L. and W. Marzocchi (2007), *Ann. Geophys.*, 50, 329–339.
- [14] McCullagh, P. and J. A. Nelder (1989), *Generalized Linear Models*, Chapman & Hall, New York.
- [15] Vermeesch, P. (2005), Statistical uncertainty associated with histograms in the Earth sciences, *JGR*, 110, B02211, doi:10.1029/2004JB003479.
- [16] Johnson, N. L. and S. Kotz (1969), *Discrete distributions*, John Wiley & Sons, New York.
- [17] Green, J. and J. Round-Turner (1986), The Error in Approximating Cumulative Binomial and Poisson Probabilities, *Teaching Statistics*, 8, 53–58.
- [18] Borradaile, G. J. (2003), *Statistics of Earth Science Data: Their Distribution in Time, Space and Orientation*, Springer-Verlag, Berlin and Heidelberg.
- [19] Koravos, G. Ch., I. G. Main, T. M. Tsapanos and R. M. W. Musson (2003), Maximum earthquake magnitudes in the Aegean area constrained by tectonic moment release rates, *Geophys. J. Int.*, 152, 94–112.
- [20] Leach, C. (1979), *Introduction to Statistics: A Non-Parametric Approach for the Social Sciences*, John Wiley & Sons, New York.